

AD-A241 899



✓m  
(2)

# NAVAL POSTGRADUATE SCHOOL Monterey, California



## THESIS

A PROBABILISTIC DERIVATION  
OF  
STIRLING'S FORMULA

by

Li, Hsin-Yun

March 1991

Thesis Advisor:

C. L. Frenzen

Approved for public release; distribution is unlimited.

91-14279



91 10 28 028

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE				Form Approved OMB No 0704-0188	
1a REPORT SECURITY CLASSIFICATION Unclassified			1b RESTRICTIVE MARKINGS		
2a SECURITY CLASSIFICATION AUTHORITY			3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.		
2b DECLASSIFICATION/DOWNGRADING SCHEDULE					
4 PERFORMING ORGANIZATION REPORT NUMBER(S)			5 MONITORING ORGANIZATION REPORT NUMBER(S)		
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School		6b OFFICE SYMBOL (if applicable) MA	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School		
6c ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000			7b ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000		
8a NAME OF FUNDING/SPONSORING ORGANIZATION		8b OFFICE SYMBOL (if applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c ADDRESS (City, State, and ZIP Code)			10 SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO	PROJECT NO	TASK NO
					WORK UNIT ACCESSION NO
11 TITLE (Include Security Classification) A PROBABILISTIC DERIVATION OF STIRLING'S FORMULA					
12 PERSONAL AUTHOR(S) Li, Hsin-Yun					
13a TYPE OF REPORT Master's Thesis		13b TIME COVERED FROM _____ TO _____	14 DATE OF REPORT (Year, Month, Day) March 1991		15 PAGE COUNT 48
16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Gov.					
17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Asymptotic analysis, Probability, Combinatorial, and Calculus		
19 ABSTRACT (Continue on reverse if necessary and identify by block number) Stirling's formula is one of the most frequently used results from asymptotics. It is used in probability and statistics, algorithm analysis and physics. In this thesis we shall give a new probabilistic derivation of Stirling's formula. Our motivation comes from sampling randomly with replacement from a group of n distinct alternatives. Usually a repetition will occur before we obtain all n distinct alternatives consecutively. We shall show that Stirling's formula can be derived and interpreted as follows: as $n \rightarrow \infty$ the expected total number of distinct alternatives we must sample before all n are obtained consecutively is asymptotically equal to the expected number of attempts we make to obtain all n distinct alternatives consecutively times the expected number of distinct alternatives obtained per attempt.					
20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21 ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a NAME OF RESPONSIBLE INDIVIDUAL C.L. Frenzen			22b TELEPHONE (Include Area Code) (408) 646-2435		22c OFFICE SYMBOL MA/Fr

Approved for public release; distribution is unlimited.

A Probabilistic Derivation of Stirling's Formula

by

Li, Hsin-Yun  
LCDR, Republic of China Navy  
B.S., Chinese Naval Academy, 1980

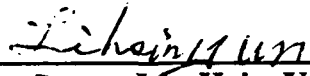
Submitted in partial fulfillment of the requirements  
for the degree of

MASTER OF SCIENCE IN APPLIED MATHEMATICS

from the

NAVAL POSTGRADUATE SCHOOL  
March 1991

Author:

  
Li, Hsin-Yun

Approved by:

  
C. L. Frenzen, Thesis Advisor



M. Aissen, Second Reader



Harold M. Fredricksen, Chairman  
Department of Mathematics

## ABSTRACT

Stirling's formula is one of the most frequently used results from asymptotics. It is used in probability and statistics, algorithm analysis and physics. In this thesis we shall give a new probabilistic derivation of Stirling's formula. Our motivation comes from sampling randomly with replacement from a group of  $n$  distinct alternatives. Usually a repetition will occur before we obtain all  $n$  distinct alternatives consecutively. We shall show that Stirling's formula can be derived and interpreted as follows : as  $n \rightarrow \infty$  the expected total number of distinct alternatives we must sample before all  $n$  are obtained consecutively is asymptotically equal to the expected number of attempts we make to obtain all  $n$  distinct alternatives consecutively times the expected number of distinct alternatives obtained per attempt.

1. AUTHOR	✓
2. TITLE	
3. SUBTITLE	
4. AUTHORING	
5. DISTRIBUTION	
6. BY	
7. PRODUCTION	
8. IS	
9. DATE	
10. SPECIAL	
A-1	



## TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	THE PROBLEM.....	1
B.	MOTIVATION.....	2
II.	COMBINATORIAL SET UP.....	3
A.	DEFINITION.....	3
B.	THE AVERAGE OR EXPECTED NUMBER OF DISTINCT BALLS BEFORE A REPETITION.....	3
C.	A RELATED GAME.....	10
D.	ASYMPTOTICS FOR THE TWO GAMES.....	16
1.	An Inequality for $E_n$ and a Partial Sum of $E_n^*$ .....	17
2.	The Divergence of $E_n$ and $E_n^*$ .....	20
3.	The Leading Asymptotic Contributions to $E_n$ and $E_n^*$ .....	23
4.	The Proof that $E_n$ is Asymptotic to $E_n^*$ .....	29
III.	STIRLING'S FORMULA.....	32
A.	THE COMMON ASYMPTOTIC VALUE OF $E_n$ AND $E_n^*$ .....	32
B.	THE ASYMPTOTICS OF THE EXPECTED TOTAL NUMBER OF DISTINCT BALLS.....	35
C.	STIRLING'S FORMULA.....	38
IV.	CONCLUSION.....	40
	LIST OF REFERENCES.....	42
	INITIAL DISTRIBUTION LIST.....	43

## L INTRODUCTION

### A. THE PROBLEM

Asymptotic analysis is important in many areas of modern science, such as the theory of probability, complex analysis and applied mathematics.

Because the factorial function and its asymptotic behavior are often needed in mathematics and engineering, Stirling's formula

$$n! \sim n^n e^{-n} \sqrt{2\pi n}, \quad (n \rightarrow \infty), \quad (1)$$

is one of the most important and frequently used asymptotic formulas. The notation in (1) means that the ratio of the left side and the right side tends to one as  $n$  tends to infinity.

There are several ways to prove Stirling's formula. For example, one can take the logarithm of  $n!$  and use Wallis's formula to obtain the factor of  $(\pi)^{1/2}$ . For this type of proof, see [Ref. 1].

Alternatively, one can start with the integral representation

$$n! = \int_0^\infty t^n e^{-t} dt$$

and use Laplace's method for integrals to evaluate it asymptotically. See [Ref. 2] for this type of approach.

All these methods use many techniques from mathematical analysis and some of them are quite sophisticated. We will pursue a new way to prove Stirling's formula using a discrete or combinatorial approach.

This method of proof was mentioned as a research problem in [Ref. 3 ], and the purpose of this thesis is to present a solution to this problem.

## B. MOTIVATION

From (1), Stirling's formula can also be written as

$$\frac{n!}{n^n} \sim \frac{\sqrt{2\pi n}}{2} \left(\frac{e}{n}\right)^{-1}, \quad (n \rightarrow \infty).$$

Imagine a box filled with  $n$  distinct balls. We shall select balls at random with replacement. The motivation for our approach comes from noting that  $n!/n^n$  is the probability of selecting  $n$  distinct balls consecutively while  $(2\pi n)^{1/2}/2$  is the asymptotic expected number of distinct balls obtained before a repetition as  $n \rightarrow \infty$ . These expressions appear in Stirling's formula as written above thus indicating that a combinatorial proof might be possible. The purpose of the next section is to define the combinatorial set up in more detail.

## II. COMBINATORIAL SET UP

### A. DEFINITION OF THE GAME

Imagine a box filled with identical balls numbered from one to  $n$ . We draw a ball from the box at random, write down its number, and replace it, mixing the balls well so that our next draw is also made at random. If we continue this process we will eventually get a repetition for there are only  $n$  distinct balls and we must certainly repeat a number by our  $(n+1)^{\text{st}}$  draw.

We are interested in the task of drawing out all  $n$  balls consecutively in this manner. We mean by this that after  $n$  consecutive draws, recording the numbers as we draw, we wish to obtain a permutation of the sequence  $(1, 2, 3, \dots, n)$ . If we obtain a repetition before the desired result, then we start over from the beginning.

The following three questions are of interest:

- What is the average or expected number  $E_n$  of distinct balls obtained before a repetition occurs?
- What is the average or expected total number  $T_n$  of distinct balls (adding up the number of distinct balls obtained in the first game, the second game ... etc.) selected before obtaining  $n$  consecutive distinct balls (i.e., some permutation of  $(1, 2, 3, \dots, n)$ ).
- What are the asymptotics of  $E_n$  and  $T_n$  as  $n \rightarrow \infty$ ?

### B. THE AVERAGE OR EXPECTED NUMBER OF DISTINCT BALLS BEFORE A REPETITION

Let  $p_j$  be the probability that we get exactly  $j$  distinct balls before a repetition. In other words, since only  $j$  distinct balls are obtained,  $p_j$  is the probability of getting a repetition on the  $(j+1)^{\text{st}}$  draw, and not before.



Since we assume each draw is independent, the probability of drawing a specific ball at any time is just  $1/n$ . We always obtain at least one distinct ball in any play of the game. To obtain exactly one distinct ball in a game we must get a repetition on our second draw. The probability of doing this is  $1/n$  and therefore the probability of obtaining only one distinct ball is

$$p_1 = \frac{1}{n}$$

The probability  $p_1$  also can be written as

$$p_1 = \frac{n-0}{n} \frac{1}{n}$$

To draw two distinct balls before a repetition we must get distinct balls on both the first and second draws. The third draw must then be a repetition of either the first or second draw. Thus we are looking for a three-tuple where the first two elements are distinct and the third element is a repetition of either the first or second element. Since all elements come from  $\{1, 2, 3, \dots, n\}$ , there are a total of  $n^3$  three-tuples, with only  $2 \cdot n \cdot (n-1)$  meeting the above requirement. Therefore,

$$p_2 = \frac{n-0}{n} \frac{n-1}{n} \frac{2}{n}$$

In the same way,

$$p_3 = \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \frac{3}{n},$$

and clearly

$$p_j = \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-j+1}{n} \frac{j}{n}, \quad (j = 1, 2, 3, \dots, n).$$

Let us check that the sum of all probabilities is one, i.e.,

$$\sum_{j=1}^n p_j = 1.$$

Since

$$\begin{aligned} p_j &= \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-j+1}{n} \frac{j}{n} \\ &= \frac{j (n) (n-1) \dots (n-j+1)}{n^{j+1}}, \end{aligned}$$

setting  $j = n$  gives

$$\begin{aligned} p_n &= \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-n+1}{n} \frac{n}{n} \\ &= \frac{n!}{n^n}. \end{aligned} \tag{2}$$

For  $j = n-1$ ,

$$\begin{aligned} p_{n-1} &= \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-(n-1)+1}{n} \frac{n-1}{n} \\ &= \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{2}{n} \frac{n-1}{n} \\ &= \frac{(n-1)(n-1)!}{n^{n-1}} \frac{n}{n} \end{aligned}$$

$$= \frac{(n-1)(n)!}{n^n},$$

and for  $j = n-2$ ,

$$\begin{aligned} p_{n-2} &= \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-(n-2)+1}{n} \frac{n-2}{n} \\ &= \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{3}{n} \frac{n-2}{n} \\ &= \frac{(n-2)(n-1)!}{n^{n-2} 2} \\ &= \frac{n(n-2)}{2} \frac{n!}{n^n}. \end{aligned}$$

In general, for  $0 \leq j \leq n-1$ ,

$$\begin{aligned} p_{n-j} &= \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} \dots \frac{n-(n-j)+1}{n} \frac{n-j}{n} \\ &= \frac{(n-j)(n-1)(n-2)\dots(j+1)}{n^{n-j}} \\ &= \frac{(n-j)(n-1)(n-2)\dots(j+1)}{n^{n-j}} \frac{n^j}{n^j} \frac{j!}{j!} \\ &= \frac{n^{j-1}(n-j)}{j!} \frac{n!}{n^n}. \end{aligned}$$

We shall now sum each  $p_j$  from  $p_n$  to  $p_1$  in reverse order beginning with

$$\begin{aligned} p_n + p_{n-1} &= [1 + (n-1)] \frac{n!}{n^n} \\ &= \frac{n n!}{n^n} \end{aligned}$$

Continuing,

$$\begin{aligned} p_n + p_{n-1} + p_{n-2} &= \frac{n n!}{n^n} + \left[ \frac{n(n-2)}{2} \right] \frac{n!}{n^n} \\ &= \left( \frac{2n + n^2 - 2n}{2} \right) \frac{n!}{n^n} \\ &= \left( \frac{n^2}{2} \right) \frac{n!}{n^n} \end{aligned}$$

and

$$\begin{aligned} p_n + p_{n-1} + p_{n-2} + p_{n-3} &= \frac{n^2}{2} \frac{n!}{n^n} + \frac{n^2(n-3)}{3!} \frac{n!}{n^n} \\ &= \left( \frac{3n^2 + n^3 - 3n^2}{6} \right) \frac{n!}{n^n} \\ &= \left( \frac{n^3}{6} \right) \frac{n!}{n^n} \end{aligned}$$

We claim that

$$p_n + p_{n-1} + \dots + p_{n-j} = \frac{n!}{n^n} \frac{n^j}{j!} \quad (0 \leq j \leq n-1).$$

We can prove this by induction. Assume

$$p_n + p_{n-1} + \dots + p_{n-k} = \frac{n!}{n^n} \frac{n^k}{k!}$$

for some  $k < n-1$ . This is the induction hypothesis. It holds for  $k = 1, 2$ , and  $3$ .

Then

$$\begin{aligned} p_n + p_{n-1} + \dots + p_{n-k} + p_{n-(k+1)} \\ &= \frac{n!}{n^n} \frac{n^k}{k!} + \frac{n^{(k+1)-1} (n-k-1)}{(k+1)!} \frac{n!}{n^n} \\ &= \frac{n!}{n^n} \frac{n^k}{k!} \left[ 1 + \frac{(n-k-1)}{(k+1)} \right] \\ &= \frac{n!}{n^n} \frac{n^k}{k!} \left( \frac{n}{k+1} \right) \\ &= \frac{n!}{n^n} \frac{n^{k+1}}{(k+1)!} \end{aligned}$$

This proves

$$\sum_{k=0}^j p_{n-k} = p_n + p_{n-1} + \dots + p_{n-j} = \frac{n!}{n^n} \frac{n^j}{j!}, \quad 0 \leq j \leq n-1. \quad (3)$$

Equation (3) now implies

$$\begin{aligned}\sum_{k=1}^n p_k &= \sum_{j=0}^{n-1} p_{n-j} \\ &= \frac{n!}{n^n} \frac{n^{n-1}}{(n-1)!} \\ &= 1.\end{aligned}$$

What is the expected number  $E_n$  of distinct balls obtained before a repetition? This expression is

$$E_n = \sum_{k=1}^n k p_k = p_1 + 2 p_2 + 3 p_3 + \dots + n p_n$$

$$\begin{aligned}&= p_1 + p_2 + p_3 + \dots + p_n \\ &\quad + p_2 + p_3 + \dots + p_n \\ &\quad \quad + p_3 + \dots + p_n \\ &\quad \quad \quad \cdot \\ &\quad \quad \quad \cdot \\ &\quad \quad \quad \cdot \\ &\quad \quad \quad \quad + p_n\end{aligned}$$

$$= 1 + (1-p_1) + (1-p_1-p_2) + \dots + (1-p_1-p_2-\dots-p_{n-1}).$$

From (3),

$$1 - p_1 - p_2 - \dots - p_j = p_{j+1} + p_{j+2} + \dots + p_n$$

$$\begin{aligned}
&= \sum_{k=0}^{n-(j+1)} p_{n-k} \\
&= \frac{n!}{n^n} \frac{n^{n-(j+1)}}{(n-(j+1))!} \\
&= \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-j}{n}, \quad (1 \leq j \leq n-1).
\end{aligned}$$

Thus

$$E_n = 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \dots + \frac{n-1}{n} \frac{n-2}{n} \dots \frac{1}{n}. \quad (4)$$

### C. A RELATED GAME

Now we consider a related game to simplify our later analysis. The rules for this new game are as follows: initially, we start with  $n$  distinct balls. Now, however, when we select a distinct ball and replace it, we also add a new ball numbered differently from all previous balls. For example, suppose we have just selected the  $k^{\text{th}}$  distinct ball. Before our next draw, we replace the  $k^{\text{th}}$  ball and add a new ball numbered  $n+k$ , so that our next draw will be from a pool of  $n+k$  equally likely distinct balls. Thus, each time we draw out a distinct ball, the number of balls in the box increases by one.

This game, like the previous one, ends when we get a repetition. However, unlike the first game, now it is possible in principle to obtain arbitrarily many distinct balls.

If the game ends, we empty the box and start over with the original  $n$  balls.

Now, what is the expected number of distinct balls  $E_n^*$  obtained when playing this second game?

Let  $p_j^*$  be the probability that we draw out  $j$  distinct balls, i.e., repetition occurs on the  $(j+1)^{\text{st}}$  draw, and not before.

Thus  $p_1^*$  is the probability that a repetition occurs on the second draw. Since we always get a distinct ball on the first draw, after the first draw, the box has  $n+1$  balls. Therefore,

$$p_1^* = \frac{n}{n} \frac{1}{n+1} = \frac{1}{n+1}.$$

In the same way,  $p_2^*$  is the probability that a repetition happens on the third draw, and not before. Therefore the first and second draws yield distinct balls so that after the first draw there are  $n+1$  balls in the box, and after the second draw there are  $n+2$  balls in the box. The probability of getting a distinct ball on the second draw is  $n/(n+1)$ , and in order to get a repetition on the third draw, we have only two choices out of  $n+2$  balls in the box. Therefore

$$p_2^* = \frac{n}{n} \frac{n}{n+1} \frac{2}{n+2}.$$

In the same way we find

$$p_3^* = \frac{n}{n} \frac{n}{n+1} \frac{n}{n+2} \frac{3}{n+3}$$



$$p_j^* = \frac{n}{n} \frac{n}{n+1} \frac{n}{n+2} \frac{n}{n+3} \cdots \frac{n}{n+j-1} \frac{j}{n+j}$$

Again we need to show that the sum of all probabilities is one. We argue as follows.

$$p_1^* = \frac{n}{n} \frac{1}{n+1}$$

$$= 1 - \frac{n}{n+1}$$

$$p_1^* + p_2^* = \left(1 - \frac{n}{n+1}\right) + \left(\frac{n}{n} \frac{n}{n+1} \frac{2}{n+2}\right)$$

$$= 1 - \frac{n}{n+1} + \frac{n}{n+1} \frac{2}{n+2}$$

$$= 1 - \left(\frac{n}{n+1}\right) \left(1 - \frac{2}{n+2}\right)$$

$$= 1 - \left(\frac{n}{n+1}\right) \left(\frac{n}{n+2}\right)$$

$$p_1^* + p_2^* + p_3^* = 1 - \frac{n}{n+1} \frac{n}{n+2} + \frac{n}{n+1} \frac{n}{n+2} \frac{3}{n+3}$$

$$= 1 - \left(\frac{n}{n+1}\right) \left(\frac{n}{n+2}\right) \left(1 - \frac{3}{n+3}\right)$$

$$= 1 - \frac{n}{n+1} \frac{n}{n+2} \frac{n}{n+3} .$$

We claim

$$\sum_{j=1}^k p_j = 1 - \frac{n^k}{(n+1)(n+2) \dots (n+k)}$$

This is true for  $k=1$ .

Assume for the induction hypothesis that

$$\sum_{j=1}^m p_j = 1 - \frac{n^m}{(n+1)(n+2) \dots (n+m)}$$

Then

$$\begin{aligned} \sum_{j=1}^m p_j + p_{m+1} &= 1 - \frac{n^m}{(n+1)(n+2) \dots (n+m)} + \left( \frac{n}{n} \frac{n}{n+1} \frac{n}{n+2} \dots \frac{n}{n+m} \frac{m+1}{n+m+1} \right) \\ &= 1 - \frac{n^m}{(n+1)(n+2) \dots (n+m)} + \frac{n^m}{(n+1)(n+2) \dots (n+m)} \frac{m+1}{n+m+1} \\ &= 1 - \left( \frac{n^m}{(n+1)(n+2) \dots (n+m)} \right) \left( 1 - \frac{m+1}{n+m+1} \right) \\ &= 1 - \frac{n^{m+1}}{(n+1)(n+2) \dots (n+m)(n+m+1)} . \end{aligned}$$

This concludes the induction proof.

Note that the expression

$$\sum_{j=1}^k p_j^* = 1 - \frac{n^k}{(n+1)(n+2)\dots(n+k)}$$

can also be written as

$$1 - \sum_{j=1}^k p_j^* = \frac{n^k}{(n+1)(n+2)\dots(n+k)} \quad (5)$$

Now let us consider

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{n^k}{(n+1)(n+2)\dots(n+k)} \\ &= \lim_{k \rightarrow \infty} \frac{1}{\left(1 + \frac{1}{n}\right) \left(1 + \frac{2}{n}\right) \dots \left(1 + \frac{k}{n}\right)} \end{aligned}$$

Since

$$1 + \frac{k}{n} \longrightarrow \infty \text{ as } k \longrightarrow \infty,$$

it follows that

$$\lim_{k \rightarrow \infty} \frac{n^k}{(n+1)(n+2)\dots(n+k)} = 0.$$

Letting  $k \rightarrow \infty$  in (5), it then follows that

$$\begin{aligned} \sum_{j=1}^{\infty} p_j^* &= \lim_{k \rightarrow \infty} \sum_{j=1}^k p_j^* \\ &= \lim_{k \rightarrow \infty} \left[ 1 - \frac{n^k}{(n+1)(n+2)\dots(n+k)} \right] \end{aligned}$$

$$= 1 - \lim_{k \rightarrow \infty} \frac{n^k}{(n+1)(n+2) \dots (n+k)}$$

$$= 1$$

Now we shall find the expected number  $E_n^*$  of distinct balls for this second game. The expected value  $E_n^*$  is

$$E_n^* = \sum_{k=1}^{\infty} k p_k^* = p_1^* + 2p_2^* + 3p_3^* + \dots$$

$$= p_1^* + p_2^* + p_3^* + p_4^* + \dots$$

$$+ p_2^* + p_3^* + p_4^* + \dots$$

$$+ p_3^* + p_4^* + \dots$$

$$+ p_4^* + \dots$$

.

.

.

$$= 1 + (1 - p_1^*) + [1 - (p_1^* + p_2^*)] + \dots + [1 - (p_1^* + p_2^* + \dots + p_j^*)] + \dots$$

$$= 1 + \frac{n}{n+1} + \frac{n}{n+1} \frac{n}{n+2} + \dots + \frac{n}{n+1} \frac{n}{n+2} \dots \frac{n}{n+j} + \dots \quad (6)$$

Note that to derive (6) we have used (5).

#### D. ASYMPTOTICS FOR THE TWO GAMES

We shall now study the expected values of the two games. No closed form expressions appear to exist for  $E_n$  and  $E_n^*$  as given in (4) and (6). However our proof of Stirling's formula, as mentioned in the introduction, requires the asymptotic behavior of  $E_n$  as  $n \rightarrow \infty$ . In this section we shall study the asymptotic behavior of  $E_n$  and  $E_n^*$ .

Recall that

$$E_n = \sum_{k=1}^n k p_k = 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \dots + \frac{n-1}{n} \frac{n-2}{n} \dots \frac{1}{n},$$

and

$$E_n^* = \sum_{k=1}^{\infty} k p_k^* = 1 + \frac{n}{n+1} + \frac{n}{n+1} \frac{n}{n+2} + \frac{n}{n+1} \frac{n}{n+2} \frac{n}{n+3} + \dots$$

Let

$$\alpha_k = \frac{n-0}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k}{n} \quad (7)$$

$$\beta_k = \frac{n}{n+0} \frac{n}{n+1} \frac{n}{n+2} \dots \frac{n}{n+k} \quad (8)$$

Then (4) becomes

$$E_n = \sum_{k=1}^n k p_k = \sum_{k=0}^{n-1} \alpha_k \quad (9)$$

while (6) becomes

$$E_n^* = \sum_{k=1}^{\infty} k p_k^* = \sum_{k=0}^{\infty} \beta_k \quad (10)$$

We are going to show that

$$\sum_{k=0}^{n-1} \alpha_k \sim \sum_{k=0}^{\infty} \beta_k \text{ as } n \rightarrow \infty$$

or

$$E_n \sim E_n^*, \quad (n \rightarrow \infty). \quad (11)$$

We shall do this in several steps which will be given in sections 1 through 4.

### 1. An Inequality for $E_n$ and a Partial Sum of $E_n^*$

The goal of this section is to obtain (15), an important inequality relating  $E_n$  to a partial sum of  $E_n^*$ .

To accomplish this, there is an inequality we shall need. It is

$$1 - x \leq e^{-x} \leq \frac{1}{1+x}, \quad x \geq 0. \quad (12)$$

We prove the inequality (12) as follows :

If  $x > 0$ , then  $-x < 0$ . By exponentiating both sides we get  $e^{-x} < 1$ , or  $1 - e^{-x} > 0$ .

If  $x < 0$ , then  $-x > 0$ . By exponentiating both sides we get  $e^{-x} > 1$ , or  $1 - e^{-x} < 0$ . Define  $I(x)$  by

$$I(x) = \int_0^x (1 - e^{-t}) dt$$

From the above inequalities, we conclude that

$$I(x) > 0, \text{ if } x \neq 0.$$

However ,

$$\begin{aligned} I(x) &= \left. t + e^{-t} \right|_0^x \\ &= x + e^{-x} - 1. \end{aligned}$$

Consequently

$$e^{-x} > 1 - x, \quad x \neq 0. \quad (13)$$

Note that equality occurs in (13) only if  $x = 0$ . When  $x \geq 0$ , this expression yields the first inequality in (12).

For the second inequality in (12), we note that (13) implies  $e^x \geq 1 + x$ , ( $x \geq 0$ ), or  $e^{-x} \leq 1/(1+x)$ , ( $x \geq 0$ ). Returning to the problem, from (7) and (12) it follows that for  $0 \leq k \leq n-1$

$$\begin{aligned} \alpha_k &= \left(1 - \frac{0}{n}\right) \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k}{n}\right) \\ &\leq e^{\frac{-0}{n}} e^{\frac{-1}{n}} e^{\frac{-2}{n}} \dots e^{\frac{-k}{n}} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{1+\frac{0}{n}} \frac{1}{1+\frac{1}{n}} \frac{1}{1+\frac{2}{n}} \cdots \frac{1}{1+\frac{k}{n}} \\
&= \frac{n}{n+1} \frac{n}{n+2} \frac{n}{n+3} \cdots \frac{n}{n+k} \\
&= \beta_k.
\end{aligned}$$

Therefore,

$$\alpha_k \leq e^{\frac{-k(k+1)}{2n}} \leq \beta_k, \quad (0 \leq k \leq n-1). \quad (14)$$

Summing the inequality in (14) we conclude that

$$\sum_{k=0}^{n-1} \alpha_k \leq \sum_{k=0}^{n-1} e^{\frac{-k(k+1)}{2n}} \leq \sum_{k=0}^{n-1} \beta_k \quad (15)$$

Note that (9), (10) and (15) imply

$$E_n^* > E_n \quad (16)$$



## 2. The Divergence of $E_n$ and $E_n^*$

We shall need the following inequality, special cases of which will be useful later. This result is given as an exercise on page 60 of [Ref. 4].

If  $\mu_i > -1$  for  $i = 1, 2, \dots, m$  where  $m > 1$  and  $\mu_1, \mu_2, \dots, \mu_m$  are all positive or negative, then

$$\prod_{i=1}^m (1 + \mu_i) > 1 + \sum_{i=1}^m \mu_i \quad (17)$$

We shall prove this by induction. For  $m = 2$  the inequality holds since

$$\begin{aligned} (1 + \mu_1)(1 + \mu_2) &= 1 + \mu_1 + \mu_2 + \mu_1\mu_2 \\ &> 1 + \mu_1 + \mu_2. \end{aligned}$$

Assume the inequality is true for  $m = k$ , i.e.,

$$\prod_{i=1}^k (1 + \mu_i) > 1 + \sum_{i=1}^k \mu_i$$

This is the induction hypothesis. Multiplying both sides by  $1 + \mu_{k+1} > 0$ , we have

$$\begin{aligned}
\prod_{i=1}^{k+1} (1 + \mu_i) &> 1 + \sum_{i=1}^{k+1} \mu_i + \mu_{k+1} \sum_{i=1}^k \mu_i \\
&> 1 + \sum_{i=1}^{k+1} \mu_i
\end{aligned}$$

The above inequality holds because all  $\mu_i$  are either positive or negative. This completes the proof.

Since

$$\alpha_k = (1 - \frac{0}{n})(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k}{n}), \quad (0 \leq k \leq n-1).$$

Clearly

$$0 < \alpha_k \leq 1, \quad (0 \leq k \leq n-1),$$

with  $\alpha_k = 1$  only when  $k = 0$ .

We claim that

$$(1 - \frac{1}{n}) \dots (1 - \frac{k}{n}) \geq 1 - (\frac{1}{n} + \dots + \frac{k}{n}), \quad (1 \leq k \leq n-1).$$

This follows from (17) with  $\mu_i = -i/n$  and  $m = n-1$ . In this case  $-1 < \mu_i < 0$  for  $i = 1, 2, \dots, n-1$ . Using the definition of  $\alpha_k$  and admitting the case  $k = 0$ , we conclude

$$1 - \left( \frac{0}{n} + \frac{1}{n} + \dots + \frac{k}{n} \right) \leq \alpha_k \leq 1, \quad (0 \leq k \leq n-1). \quad (18)$$

Since

$$\frac{0}{n} + \frac{1}{n} + \dots + \frac{k}{n} = \frac{k(k+1)}{2n},$$

(18) may be written as

$$1 - \frac{k(k+1)}{2n} \leq \alpha_k \leq 1, \quad (0 \leq k \leq n-1).$$

It follows that if  $k^2 = o(n)$  as  $n \rightarrow \infty$  (i.e.,  $k^2/n \rightarrow 0$  as  $n \rightarrow \infty$ ), then

$$\alpha_k \rightarrow 1 \text{ as } n \rightarrow \infty.$$

So, for example, when  $k \leq n^{1/4}$

$$\alpha_k \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This implies  $E_n \rightarrow \infty$ , as  $n \rightarrow \infty$ , and since

$$E_n^* > E_n$$

by (16), it follows that  $E_n^* \rightarrow \infty$  as  $n \rightarrow \infty$  too.

Thus, in both games the expected number of distinct balls obtained before a repetition occurs tends to infinity as the initial number of balls in the box tends to infinity.

### 3. The Leading Asymptotic Contributions to $E_n$ and $E_n^*$

In this section we shall show that

$$E_n \sim \sum_{k=0}^{k_n} \alpha_k, \quad E_n^* \sim \sum_{k=0}^{k_n} \beta_k, \quad (n \rightarrow \infty),$$

when

$$k_n = \left\lfloor n^{\frac{1}{2} + \varepsilon} \right\rfloor, \quad (0 < \varepsilon < \frac{1}{6}).$$

This result determines the leading asymptotic contribution to  $E_n$  and  $E_n^*$ .

In order to show this, we shall need Bernoulli's inequality:

$$(1+x)^m > 1+mx, \text{ if } m > 1, x > -1 \text{ and } x \neq 0. \quad (19)$$

Bernoulli's inequality is a special case of the inequality in (17). Choose  $\mu_i = x$  for  $i = 1, 2, \dots, m$  when  $x \neq 0$  and  $x > -1$ . Then (17) implies (19).

Using (19) for any positive integer  $k$  with  $2 \leq k < n$ ,  $(1 - k/n) < (1 - 1/n)^k$ , and  $(1 - k/n) = (1 - 1/n)^k$ , for  $k = 1$  or  $k = 0$ .

Now we define

$$k_n = \left\lfloor n^{\frac{1}{2} + \epsilon} \right\rfloor, \quad (0 < \epsilon < \frac{1}{6}). \quad (20)$$

The number  $k_n$  is the largest integer less than or equal to  $n^{\frac{1}{2} + \epsilon}$ , i.e.,

$$(n^{\frac{1}{2} + \epsilon} - 1) < k_n \leq n^{\frac{1}{2} + \epsilon}.$$

If  $k > k_n$  and  $k$  is an integer then

$$n^{\frac{1}{2} + \epsilon} < k_n + 1 \leq k$$

so that

$$\frac{k^2}{n} > n^{2\epsilon}, \text{ i.e., } \frac{-k^2}{n} < -n^{2\epsilon}. \quad (21)$$

Recall that

$$E_n = \sum_{k=0}^{n-1} \beta_k + \sum_{k=n}^{\infty} \beta_k.$$

If  $k \geq n$ ,

$$\beta_k = \frac{n}{n+0} \frac{n}{n+1} \frac{n}{n+2} \cdots \frac{n}{n+n} \frac{n}{2n+1} \cdots \frac{n}{2n+k-n},$$

so that

$$\beta_k \leq \left(\frac{1}{2}\right)^{k-n}.$$

Consequently,

$$\begin{aligned}
\sum_{k=n}^{\infty} \beta_k &\leq \sum_{k=n}^{\infty} 2^{n-k} \\
&= \sum_{j=0}^{\infty} 2^{-j} \\
&= 1 + \frac{1}{2} + \frac{1}{2^2} + \dots \\
&= \frac{1}{1 - \frac{1}{2}} \\
&= 2.
\end{aligned} \tag{22}$$

Since  $E_n^*$  diverges, it follows that  $E_n^* \sim \sum_{k=0}^{n-1} \beta_k$ , ( $n \rightarrow \infty$ ).

Now, for  $k \leq n$ ,

$$\begin{aligned}
\beta_k &= \frac{n}{n+0} \frac{n}{n+1} \frac{n}{n+2} \dots \frac{n}{n+k} \\
&= \left(1 - \frac{0}{n+0}\right) \left(1 - \frac{1}{n+1}\right) \left(1 - \frac{2}{n+2}\right) \dots \left(1 - \frac{k}{n+k}\right) \\
&\leq \left(1 - \frac{0}{2n}\right) \left(1 - \frac{1}{2n}\right) \left(1 - \frac{2}{2n}\right) \dots \left(1 - \frac{k}{2n}\right) \\
&\leq \left(1 - \frac{1}{2n}\right)^0 \left(1 - \frac{1}{2n}\right)^1 \left(1 - \frac{1}{2n}\right)^2 \dots \left(1 - \frac{1}{2n}\right)^k
\end{aligned}$$

by (19). Since the right side of the above inequality can be written as

$$\left(1 - \frac{1}{2n}\right)^{0+1+2+\dots+k},$$

and  $0 + 1 + 2 + \dots + k = k(k+1)/2$ , it follows that

$$\beta_k \leq \left(1 - \frac{1}{2n}\right)^{\frac{k(k+1)}{2}}.$$

This inequality implies

$$\begin{aligned} \beta_k &\leq \left(1 - \frac{1}{2n}\right)^{\frac{k^2}{2}} \\ &\leq \left(e^{-\frac{1}{2n}}\right)^{\frac{k^2}{2}}, \quad (\text{by (12)}) \\ &= e^{-\frac{k^2}{4n}}. \end{aligned}$$

Since (21) implies

$$\frac{-k^2}{4n} < \frac{-n^2 \varepsilon}{4}.$$

for  $k_n < k < n$ , it follows that

$$\beta_k < e^{-\frac{n^2 \varepsilon}{4}}, \quad (k_n < k < n).$$

Now recall from (14), that when  $k \leq n-1$

$$\alpha_k \leq \beta_k.$$

Combining the last two inequalities then yields

$$\alpha_k \leq \beta_k < e^{\frac{-n^{2k}}{4}}, \quad (k_n < k < n).$$

It then follows that

$$\sum_{k=k_n+1}^{n-1} \alpha_k \leq \sum_{k=k_n+1}^{n-1} \beta_k < n e^{\frac{-n^{2k}}{4}}.$$

Since  $n e^{\frac{-n^{2k}}{4}} \rightarrow 0$  as  $n \rightarrow \infty$ , it follows that

$$\sum_{k=k_n+1}^{n-1} \alpha_k \rightarrow 0, \quad (23)$$

$$\sum_{k=k_n+1}^{n-1} \beta_k \rightarrow 0 \quad (24)$$

as  $n \rightarrow \infty$ .

Recall that

$$E_n = \sum_{k=0}^{\infty} \beta_k$$



$$= \sum_{k=0}^{k_n} \beta_k + \sum_{k=k_n+1}^{n-1} \beta_k + \sum_{k=n}^{\infty} \beta_k$$

Since  $E_n \rightarrow \infty$  as  $n \rightarrow \infty$  and the second and third sums in the above expression are finite by (22) and (24). It follows that

$$E_n \sim \sum_{k=0}^{k_n} \beta_k, \quad (n \rightarrow \infty). \quad (25)$$

Similarly,

$$E_n = \sum_{k=0}^{k_n} \alpha_k + \sum_{k=k_n+1}^{n-1} \alpha_k.$$

Since  $E_n$  also diverges as  $n \rightarrow \infty$ , and the second sum above is finite by (23), as  $n \rightarrow \infty$  we conclude

$$E_n \sim \sum_{k=0}^{k_n} \alpha_k, \quad (n \rightarrow \infty). \quad (26)$$

#### 4. The Proof that $E_n$ is Asymptotic to $E_n^*$

From the above analysis, we conclude that

$$E_n \sim \sum_{k=0}^{k_n} \alpha_k, \quad E_n^* \sim \sum_{k=0}^{k_n} \beta_k, \quad (n \rightarrow \infty).$$

We shall now show that

$$\sum_{k=0}^{k_n} \alpha_k \sim \sum_{k=0}^{k_n} \beta_k, \quad (n \rightarrow \infty),$$

or from (9) and (10) that

$$E_n \sim E_n^* \text{ as } n \rightarrow \infty,$$

thus proving (11).

For  $0 \leq k \leq k_n$ ,

$$\begin{aligned} \frac{\alpha_k}{\beta_k} &= \frac{(1 - \frac{1}{n})(1 - \frac{2}{n}) \dots (1 - \frac{k}{n})}{(\frac{n}{n+1})(\frac{n}{n+2}) \dots (\frac{n}{n+k})} \\ &= \frac{(\frac{n-1}{n})(\frac{n-2}{n}) \dots (\frac{n-k}{n})}{(\frac{n}{n+1})(\frac{n}{n+2}) \dots (\frac{n}{n+k})} \\ &= (\frac{n^2-1^2}{n^2})(\frac{n^2-2^2}{n^2}) \dots (\frac{n^2-k^2}{n^2}) \end{aligned}$$

$$\begin{aligned}
&= \left(1 - \left(\frac{1}{n}\right)^2\right) \left(1 - \left(\frac{2}{n}\right)^2\right) \dots \left(1 - \left(\frac{k}{n}\right)^2\right) \\
&\geq \left(1 - \left(\frac{k}{n}\right)^2\right)^k \\
&\geq 1 - \frac{k^3}{n^2} \quad (\text{by Bernoulli's inequality}) \\
&\geq 1 - \frac{k_n^3}{n^2}.
\end{aligned}$$

From this analysis and (14), it follows that

$$\beta_k \left(1 - \frac{k_n^3}{n^2}\right) \leq \alpha_k \leq \beta_k, \quad (0 \leq k \leq k_n).$$

Summing the above inequality gives

$$\left(1 - \frac{k_n^3}{n^2}\right) \sum_{k=0}^{k_n} \beta_k \leq \sum_{k=0}^{k_n} \alpha_k \leq \sum_{k=0}^{k_n} \beta_k \quad (27)$$

By the definition in (20),

$$k_n = \left\lfloor n^{\frac{1}{2} + \varepsilon} \right\rfloor, \quad (0 < \varepsilon < \frac{1}{6}),$$

so

$$\frac{k_n^3}{n^2} \leq \frac{(n^{\frac{1}{2}+\epsilon})^3}{n^2} \leq \frac{n^{\frac{3}{2}+3\epsilon}}{n^2} = n^{\frac{1}{2}+3\epsilon}$$

Since  $0 < \epsilon < 1/6$ , it follows that

$$\frac{k_n^3}{n^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Letting  $n \rightarrow \infty$  in (27), we can conclude that

$$\sum_{k=0}^{k_n} \alpha_k \sim \sum_{k=0}^{k_n} \beta_k, \quad (n \rightarrow \infty) \quad (28)$$

From (25) and (26), it then follows that

$$E_n \sim E_n^*, \quad (n \rightarrow \infty).$$

### III. STIRLING'S FORMULA

#### A. THE COMMON ASYMPTOTIC VALUE OF $E_n$ AND $E_n^*$

In Chapter II we proved that

$$E_n \sim E_n^*, \quad (n \rightarrow \infty).$$

In this chapter we will determine the common asymptotic value of  $E_n$  and  $E_n^*$ .

From (14) we have

$$\alpha_k \leq e^{\frac{-k(k+1)}{2n}} \leq \beta_k, \quad (0 \leq k \leq n-1).$$

Hence,

$$\sum_{k=0}^{k_n} \alpha_k \leq \sum_{k=0}^{k_n} e^{\frac{-k(k+1)}{2n}} \leq \sum_{k=0}^{k_n} \beta_k \quad (29)$$

Letting  $n \rightarrow \infty$  in (29), and using (28) we conclude

$$\sum_{k=0}^{k_n} \alpha_k \sim \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}} e^{\frac{-k}{2n}}, \quad (n \rightarrow \infty), \quad (30)$$

$$\sum_{k=0}^{k_n} \beta_k \sim \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}} e^{\frac{-k}{2n}}, \quad (n \rightarrow \infty).$$

Consider the following inequality:

$$e^{\frac{-k_n}{2n}} \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}} \leq \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}} e^{\frac{-k}{2n}} \leq \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}}. \quad (31)$$

Since  $k_n/n = o(1)$  as  $n \rightarrow \infty$ , we conclude from (31) that

$$\sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}} e^{\frac{-k}{2n}} \sim \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}}, \quad (n \rightarrow \infty).$$

From (25), (26) and (30) it then follows that

$$E_n \sim \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}}, \quad (n \rightarrow \infty), \quad (32)$$

and

$$E_n^* \sim \sum_{k=0}^{k_n} e^{\frac{-k^2}{2n}}, \quad (n \rightarrow \infty). \quad (33)$$

To asymptotically estimate the sum in (32) and (33), note the function

$$f(x) = e^{-x^2}$$

is positive and monotone decreasing on  $[0, \infty)$ . Setting

$$h = \frac{1}{\sqrt{2n}},$$

we then have

$$f(kh) = e^{-\left(\frac{k}{\sqrt{2n}}\right)^2} = e^{-\frac{k^2}{2n}},$$

and

$$\int_0^{(k_n+1)h} f(x) dx \leq \sum_{k=1}^{k_n} h f(kh) \leq \int_0^{k_n h} f(x) dx. \quad (34)$$

Since

$$k_n > n^{\frac{1}{2}+\epsilon} - 1,$$

and  $k_n h = \frac{k_n}{\sqrt{2n}}$ , it follows that

$$k_n h > \frac{1}{\sqrt{2n}} (n^{\frac{1}{2}+\epsilon} - 1),$$

or

$$k_n h > \frac{1}{\sqrt{2}} \left( n^{\epsilon} - \frac{1}{\sqrt{n}} \right).$$

Hence,  $k_n h \rightarrow \infty$  and  $h \rightarrow 0$ , as  $n \rightarrow \infty$ .

Letting  $n \rightarrow \infty$  in (34), we conclude

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{\sqrt{2n}} \sum_{k=1}^{k_n} e^{-\frac{k^2}{2n}} \right] = \int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2},$$

or

$$\sum_{k=1}^{k_n} e^{-\frac{k^2}{2n}} \sim \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty). \quad (35)$$

Equation (32), (33) and (35) then imply

$$E_n \sim \frac{\sqrt{2\pi n}}{2}, \quad E_n^* \sim \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty). \quad (36)$$

## B. THE ASYMPTOTICS OF THE EXPECTED TOTAL NUMBER OF DISTINCT BALLS

In the first game considered above we defined  $T_n$  to be the expected total number of distinct balls obtained before winning. Let  $p$  be the probability of winning the game, and  $q = 1 - p$  be the probability of losing. Recall that one wins the game if  $n$  distinct balls are obtained consecutively and loses otherwise.

Let the random variable  $X$  be the number of times we play the game before winning. Then

$$P(X=k) = pq^{k-1}, \quad (k = 1, 2, 3, \dots).$$

The probability  $p$  of winning the game, i.e., of getting exactly  $n$  distinct balls before a repetition, is  $p = p_n = n!/n^n$ , by (2).

The expected value of  $X$  is

$$\sum_{k=1}^{\infty} kP(X=k) = \sum_{k=1}^{\infty} kpq^{k-1}$$



$$= \frac{1}{p}$$

$$= \frac{n^n}{n!}$$

This means that, on average, we must play  $n^n/n!$  times in order to win. Therefore, the expected total number of distinct balls  $T_n$  obtained before a win can be represented as follows :

$$T_n = n + \left(\frac{1}{p} - 1\right) E_n .$$

This is the value of  $T_n$  because the expected number of plays is  $1/p$  ( $=n^n/n!$ ). Among these plays, one must be a win, and furthermore that win must occur on the very last play. All the other plays are losses. When we win, we draw out  $n$  distinct balls, and when we lose, the expected number of distinct balls is  $E_n$ . Therefore,

$$T_n = n + \left(\frac{1}{p} - 1\right) E_n . \tag{37}$$

This expression can be rearranged as

$$T_n = n - E_n + \frac{E_n}{p} .$$

From (36),

$$E_n \sim \frac{\sqrt{2\pi n}}{2}, \quad E_n^* \sim \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty)$$

So  $n - E_n = O(n)$ ,  $(n \rightarrow \infty)$ . It follows that

$$T_n = O(n) + \frac{E_n}{p}. \quad (38)$$

Now consider  $E_n/p$ . Since

$$E_n \sim E_n^*, \quad (n \rightarrow \infty),$$

we can write

$$\frac{E_n}{p} \sim \frac{1}{2} \left( \frac{E_n}{p} + \frac{E_n^*}{p} \right), \quad (n \rightarrow \infty) \quad (39)$$

But

$$\begin{aligned} \frac{E_n}{p} + \frac{E_n^*}{p} &= \frac{1}{p} (E_n + E_n^*) \\ &= \frac{n^n}{n!} \left( \sum_{k=0}^{n-1} \alpha_k + \sum_{k=0}^{\infty} \beta_k \right) \\ &= \left( \frac{n^n}{n!} \right) \left[ \left( 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \dots + \frac{n-1}{n} \frac{n-2}{n} \dots \frac{1}{n} \right) + \left( 1 + \frac{n}{n+1} + \frac{n}{n+1} \frac{n}{n+2} + \dots \right) \right] \end{aligned}$$

$$= \left( \frac{n^{n-1}}{(n-1)!} + \frac{n^{n-2}}{(n-2)!} + \frac{n^{n-3}}{(n-3)!} + \dots + 1 \right) + \left( \frac{n^n}{n!} + \frac{n^{n+1}}{(n+1)!} + \frac{n^{n+2}}{(n+2)!} + \dots \right),$$

and the last expression, after rearranging, can be written as

$$\begin{aligned} & \left( 1 + n + \frac{n^2}{2!} + \dots + \frac{n^{n-2}}{(n-2)!} + \frac{n^{n-1}}{(n-1)!} \right) + \left( \frac{n^n}{n!} + \frac{n^{n+1}}{(n+1)!} + \frac{n^{n+2}}{(n+2)!} + \dots \right) \\ &= e^n. \end{aligned}$$

It follows from (39) that

$$\frac{E_n}{p} \sim \frac{e^n}{2}, \quad (n \rightarrow \infty). \quad (40)$$

With this information, letting  $n \rightarrow \infty$  in (38) yields

$$T_n \sim \frac{e^n}{2}, \quad (n \rightarrow \infty). \quad (41)$$

### C STIRLINGS' FORMULA

From (38), (40) and (41)

$$T_n \sim \frac{E_n}{p}, \quad (n \rightarrow \infty) \quad (42)$$

i.e.,

$$T_n \sim \frac{n^n}{n!} E_n, \quad (n \rightarrow \infty), \quad (43)$$

Since

$$E_n \sim \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty)$$

(43) implies

$$T_n \sim \frac{n^n}{n!} \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty).$$

Using (41), we get

$$\frac{e^n}{2} \sim \frac{n^n}{n!} \frac{\sqrt{2\pi n}}{2}, \quad (n \rightarrow \infty), \quad (44)$$

or alternatively,

$$n! \sim \frac{n^n}{e^n} \sqrt{2\pi n}, \quad (n \rightarrow \infty).$$

This can also be written as

$$n! \sim n^n e^{-n} \sqrt{2\pi n}, \quad (n \rightarrow \infty),$$

and this is Stirling's formula (1).

#### IV. CONCLUSION

We have given a new combinatorial or probabilistic derivation of Stirling's formula. Our derivation also gives another way to interpret Stirling's formula. To see this, let us consider a specific value for  $n$ . Suppose  $n = 20$ , i.e., there are 20 distinct balls in the box. The expected number of distinct balls  $E_{20}$  obtained is

$$\begin{aligned} E_{20} &= 1 + \frac{19}{20} + \frac{19}{20} \frac{18}{20} + \frac{19}{20} \frac{18}{20} \frac{17}{20} + \dots + \frac{19!}{20^{19}} \\ &= 5.293584585. \end{aligned}$$

This is fairly close to the asymptotic value of  $E_n$  as  $n \rightarrow \infty$ , i.e.,

$$\frac{\sqrt{2\pi(20)}}{2} = 5.604991216.$$

The probability of winning a game when  $n = 20$  is

$$p = p_{20} = \frac{20!}{20^{20}}.$$

so that the expected number of plays before a win is

$$\frac{1}{p} = \frac{20^{20}}{20!} \approx 43099804.$$

By (37), the expected total number of distinct balls  $T_{20}$  obtained before a win is

$$\begin{aligned} T_{20} &= 20 + (43099804 - 1)(E_{20}) \\ &= 228152473. \end{aligned} \tag{45}$$

The asymptotic formula (41) for  $T_n$  gives, for  $n = 20$ ,

$$\begin{aligned} T_n &\approx \frac{e^{20}}{2} \\ &\approx 242582598, \end{aligned}$$

and this is of the same order as (45). As  $n$  gets larger, the agreement between the exact formula for  $T_n$  and the asymptotic formula will get increasingly better. Notice that when  $n = 20$  the expected number of plays before a win is quite large.

To conclude, we can interpret Stirling's formula as written in (44) in the following way using (42). As  $n \rightarrow \infty$  the expected total number of distinct balls obtained before a win is asymptotic to the expected number of plays necessary to win times the expected number of distinct balls per play.

## LIST OF REFERENCES

1. Franklin, Philip, *Methods of Advanced Calculus* , McGraw-Hill Book Company, Inc., 1966.
2. Olver, F. W. J., *Asymptotics and Special Functions* , Academic Press Inc., 1974.
3. Graham, R.L., Knuth, D. E. and Patashnik, O., *Concrete Mathematics* , Addison Wesley, Reading, MA., 1989.
4. Hardy, G., Littlewood, J. E. and Pólya, G., *Inequalities* , Cambridge University Press, 1989.